From:	Wyatt, Robert
To:	Eric Blischke/R10/USEPA/US@EPA; ricka@bes.ci.portland.or.us; Jim.McKenna@portofportland.com
Cc:	johnt@windwardenv.com; Helle B. Andersen; Chip Humphrey/R10/USEPA/US@EPA; kpine@anchorenv.com; jworonets@anchorenv.com
Subject:	RE: Final Bioassay Evaluation Proposal
Date:	07/24/2008 05:45 PM

Hi Eric,

(b) (6) provide a technical response when

Thank you,

Bob

-----Original Message-----
From: Blischke.Eric@epamail.epa.gov
[mailto:Blischke.Eric@epamail.epa.gov]
Sent: Thursday, July 24, 2008 11:12 AM
To: Wyatt, Robert; ricka@bes.ci.portland.or.us;
Jim.McKenna@portofportland.com
Cc: johnt@windwardenv.com; Helle B. Andersen;
Humphrey.Chip@epamail.epa.gov; kpine@anchorenv.com
Subject: Final Bioassay Evaluation Proposal

Bob:

As we discussed last week, EPA has adjusted the previous proposal for the evaluation of sediment bioassays. The proposals outlined below represents EPA's final position on the bioassay evaluation. In developing this approach, EPA has considered concerns raised by your technical representatives as well as EPA's government team partners. EPA and the LWG have been discussing this issue since 2004. Recently, we have been discussing this topic in response to the February 15, 2008 Problem Formulation for the baseline ecological risk assessment developed by EPA.

Background:

Benthic Interpretive Report:

On March 17, 2006, the Lower Willamette Group submitted the Interpretive Report: Estimating Risks to Benthic Organisms using Predictive Models Based on Sediment Toxicity Tests. This report presented an evaluation of the floating percentile and logistic regression models as well as a comparison to existing sediment quality values. The stated goal of the predictive model is "to derive SQVs that are sufficiently reliable for predicting benthic toxicity within the study area" and to develop a line of evidence "for identifying areas where chemical concentrations in sediment may pose a risk to benthic invertebrates."

On July 6, 2006, EPA commented on the Benthic Interpretive Approach. The LWG responded to these comments on September 1, 2006. In the LWG response to comments, there were a number of comments that the LWG identified as category 1 - strongly disagree; cannot accept. In particular, the LWG disagreed with EPA's comment to include the Hyalella growth endpoint in the floating percentile model and to consider effects level 1 in the development of the predictive models. In addition, the LWG agreed to the use of the alternative logistic regression model using a larger, non-site specific, freshwater database for the Hyalella 28-day growth and survival test as a complimentary line of evidence to the floating percentile model. The LWG also agreed to use the revised logistic regression model based on the Hyalella pooled endpoint and the floating percentile model based on Chironomus growth, Chironomus mortality and Hyalella morality endpoints as separate lines of evidence in assessing risks to the benthic community.

Round 2 Report:

On February 21, 2007, the LWG submitted the Comprehensive Round 2 Site Characterization Summary and Data Gaps Report. In the Round 2 Report, the evaluation of benthic risks considered the floating percentile model - effect levels 2 and 3 for the Chironomus growth, Chironomus mortality and Hyalella morality endpoints and the logistic regression model at the effect level 2 for the pooled Hyalella and Chironomus endpoints. Although the Round 2 report utilized the logistic regression model for the identification of Round 2 Chemicals of Potential Concern (COPCs; see Table 9.3-1 of the Round 2 Report), the logistic regression model was not used to develop initial areas of potential concern (iAOPCs) due to the following concerns: Irreproducibiliy of the logistic regression model; the predictive ability of the Hyalella growth endpoint, and the reduction in predictive accuracy when combining the two models.

EPA considered the logistic regression model and the Hyalella growth endpoint in our evaluation of benthic risks for the purposed of identifying Round 3B data gaps. However, during the finalization of the field sampling plan for sediment toxicity testing, EPA and the LWG could not reach agreement on the use of the Hyalella growth endpoint in the application of the predictive models and instead agreed to identify sediment sampling locations, in part, based on an evaluation of the empirical Hyalella growth toxicity testing.

BERA Problem Formulation

On February 15, 2008, EPA submitted the Problem Formulation for the Baseline Ecological Risk Assessment to the LWG. The purpose of the problem formulation was to guide the development of the baseline ecological risk assessment. The problem formulation required evaluation of the empirical toxicity results at the 10, 20 and 30 % difference from control level and the floating percentile model at the 20% and 30% effect level. In addition, the problem formulation required a substitution of the Hyalella growth endpoint with a total biomass endpoint. suggested pooling of endpoints to improve model performance, recommended incorporation of the Round 3 Data into the models and recommended reconciling the chemicals evaluated in the two models to the extent possible.

Current Status

Post Problem Formulation Discussions:

Following submittal of the problem formulation by EPA, a series of discussions took place in an effort to resolve discrepancies between the Round 2 Report, the Problem Formulation and previously submitted documents such as the benthic interpretation report and the 2005 Technical Memorandum - Estimating Risks to the Benthic Community using Sediment Toxicity Tests. A number of approaches were considered including adjusting the effect levels for the Hyalella growth endpoint and incorporation of the RSET one-hit/two-hit approach into the floating percentile model.

Our most recent discussion took place on Friday, July 11, 2008. Burt Shephard and I spoke with John Toll and Helle Anderson about the evaluation of benthic risk. At the end of the discussion, we came up with the following approach:

1) Evaluate the empirical toxicity data as we have described - a hit is a statistically significant difference from control for any of the four endpoints.
2) Substitute total biomass for the growth endpoint for both the Hyalella and the chironomus tests.
3) Empirical data will be further refined by classifying the toxicity tests into minor (10%) moderate (20%) and severe effects (30%).
4) For the LRM and FPM, we will pool the growth (biomass) and mortality endpoints for chironomus and again for Hyalella.
5) Pooling will be based on use of the most sensitive endpoint (growth or mortality) resulting in two LRM and two FPM models.
6) The evaluation of the bioassay data for the development of the predictive models will be based on the following hit thresholds:
   - Chironomus Growth - 30%
   - Chironomus Mortality - 20%
   - Hyalella Growth - 40%
   - Hyalella Mortality - 20%
7) These thresholds will apply to both the logistic and floating percentile models.
8) The results from these models will be equivalent to site specific probable effect levels.
9) The draft RI report will present an evaluation of the hit thresholds used in the predictive models. The evaluation will compare the separation of sediment chemistry distributions at the hit and no hit stations as a way to assess the utility of using lower hit thresholds in the predictive models, evaluate the reliability of the predictive models and make recommendations regarding the optimization of model performance.
10) The model results will be used in the conjunction of other lines of evidence in the baseline risk assessment and in the development of PRGs.

Evaluation:

Subsequent discussions with our project team raised concerns about the thresholds for the floating percentile model evaluation. There was a strong sense that two thresholds should be evaluated and that the 40% threshold for the Hyalella growth endpoint was too high. In response to these concerns, EPA further evaluated information presented in the March 17, 2006 benthic interpretation report. The review focused on the floating percentile model and considered both the reliability parameters presented in Table 5-3 of the report as well as an evaluation of the differences between the hit and no-hit distributions as presented in Appendix D of the report.

Based on this evaluation, for the Chironomus growth and mortality and Hyalella mortality endpoints, the best performers are the 20% Chironomus growth and 30% Hyalella mortality and Chironomus mortality endpoints. However, the difference between the 20% and 30% effect thresholds is slight. The 10% effect threshold for these three endpoints show reduced performance based on reliability and difference between hit and no-hit distributions.

For the Hyalella growth endpoint, EPA acknowledges the high incidence of false positives at all three effect levels. In addition, EPA acknowledges that the difference between the hit and no-hit distributions are more difficult to discern than the other three endpoints. That said, the information presented in the reports suggests that a floating point model can be developed for the Hyalella growth endpoint. Further, evaluation of the empirical Hyalella growth data suggests that Hyalella growth at the 10% and 20% difference from control hit thresholds can be used to delineate the extent of contamination at the Portland Harbor Site.

Evaluation of the hit/no-hit distributions demonstrates 10% effect level has the greatest difference between the hit and no-hit distributions and has a reasonably good predicted hit reliability. However, it still suffers from a false positive rate of greater than 50%. The predicted hit reliability drops significantly at the 20% effect level and is only 27% at the 30% effect level. EPA is currently unable to evaluate the

LWG proposed 40% effect level for the Hyalella growth endpoint.

Given the uncertainty surrounding the utility of applying the floating percentile model to the Hyalella growth endpoint, EPA believes it is prudent to evaluate two different thresholds. Based on the evaluation summarized above, the 10% effect level seems to perform better than the 20% and 30% effect levels. Evaluation of both the 10% and 40% effect level will allow us to bracket the range of Hyalella growth effect levels considered and should facilitate evaluation and optimization of the model performance.

EPA acknowledges that due to the large number of sources and source types at the Portland Harbor site, the predictive model results do not necessarily match up well with the empirical bioassay results. As a result, the necessary analysis must be performed in the baseline risk assessment to determine the optimum hit threshold or thresholds. These results will be used along with other lines of evidence (e.g., SQGs, application of benthic tissue TRVs and BSAFs) to identify areas that pose risk to the benthic community and develop sediment cleanup levels protective of the benthic community.

Final Proposal:

Based on the evaluation outlined above, EPA is prepared to direct the LWG to evaluate benthic risks according to the following:

1) Evaluate the empirical toxicity data as we have described - a hit is a statistically significant difference from control for any of the four endpoints.
2) Substitute total biomass for the growth endpoint for both the Hyalella and the chironomus tests.
3) Empirical data will be further refined by classifying the toxicity tests into minor (10%) moderate (20%) and severe effects (30%).
4) For the Logistic Regression Model, the development of the predictive models will be based on the pooled Hyalella and pooled Chironomus endpoints at the 20% effect level.
5) Adjustment of the probability of toxicity (Pr) used to distinguish no effects, minor effects, moderate effects and severe effects from the current 40% and 60% may be considered.
6) Round 3B sediment toxicity data should be incorporated into the two predictive models.
7) EPA will make the non-Portland Harbor site data available to the LWG for evaluation of the alternative logistic regression model developed by NOAA (pooled Hyalella Growth endpoint only).
8) For the Floating Point Model, the development of the predictive models will be based on the following hit thresholds:
   - Chironomus Growth - 20% and 30%
   - Chironomus Mortality - 20% and 30%
   - Hyalella Growth - 10% and 40%
   - Hyalella Mortality - 20% and 30%
9) The draft RI report will present an evaluation of the hit thresholds used in the predictive models. The evaluation will compare the separation of sediment chemistry distributions at the hit and no hit stations as a way to assess the hit thresholds in the predictive models for possible adjustment, evaluate the reliability of the predictive models and make recommendations regarding the optimization of model performance.
10) The model results will be used in the conjunction of other lines of evidence in the baseline risk assessment and in the development of PRGs.

EPA believes that the above approach is consistent with approaches outlined in the March 2006 Benthic Interpretation Report, the Round 2 Report and EPA's Problem Formulation. Further, the approach incorporates EPA's long-standing desire to incorporate the Hyalella Growth endpoint into the floating percentile model while at the same time addressing the LWG's desire to evaluate Hyalella growth at a higher effect level (40%).

If you have any questions regarding the approach outlined above, please contact me.

Thanks, Eric